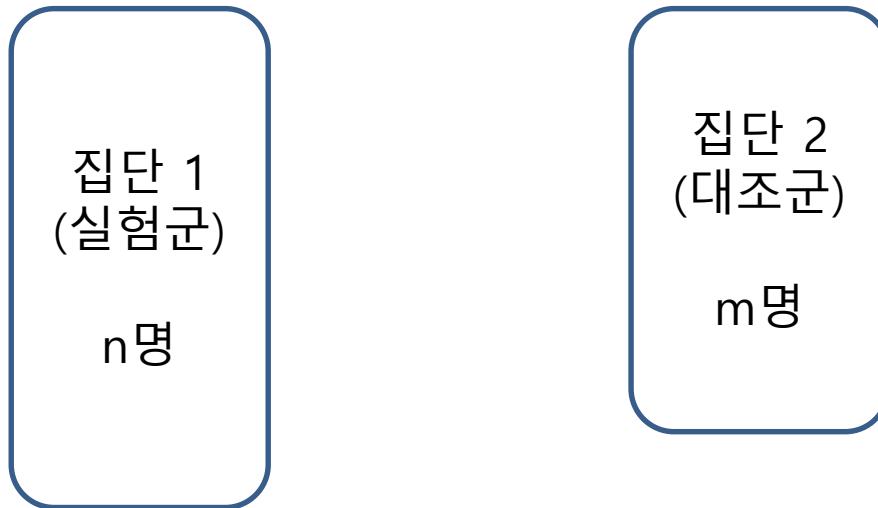


# 조사 자료의 분석과 해석 방법 : 비교 분석을 중심으로

# 독립집단 T-검정

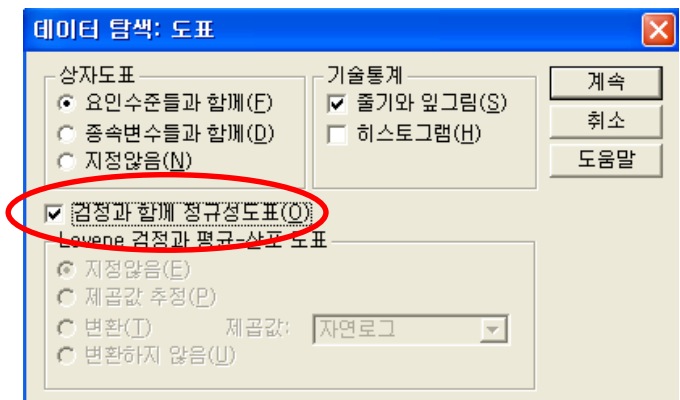
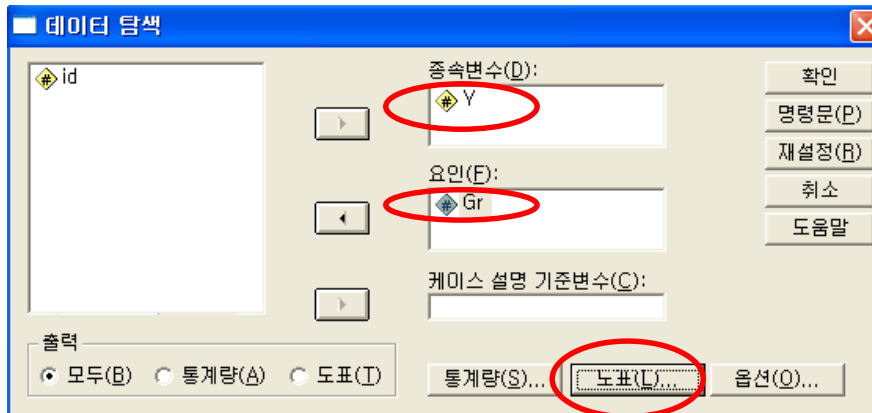
- 의미 : 두 집단의 평균 차이를 검정하는 방법
- 연구설계 : 자료수집



## – 가정

- 1) 두 집단은 서로 독립 : 두 집단은 관계가 없다는 의미
- 2) 연구 관심변수(Y)는 수치자료로 측정
- 3) 측정 자료는 정규분포로 가정 : 실제 연구에서는 분석을 통해 확인해야 함

# SPSS에서의 정규성 검정



정규성 검정

	Gr	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		통계량	자유도	유의확률	통계량	자유도	유의확률
Y	1.00	.135	7	.200*	.976	7	.936
	2.00	.163	7	.200*	.948	7	.715
	3.00	.128	7	.200*	.988	7	.989

\*, 이것은 참인 유의확률의 하한값입니다.

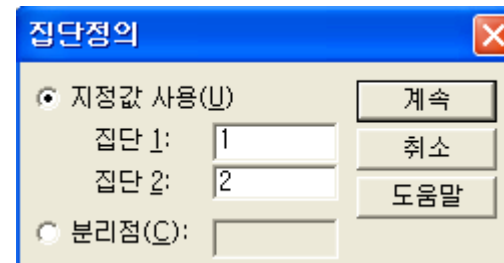
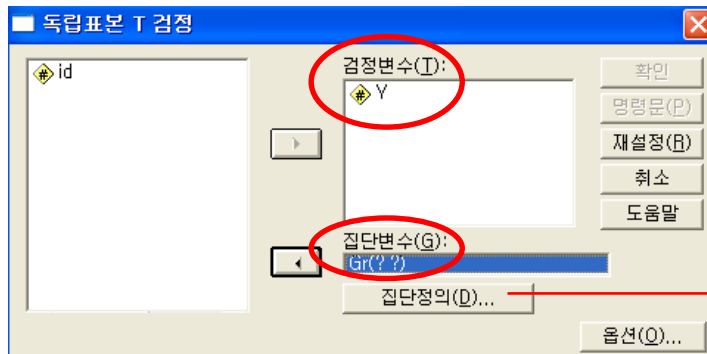
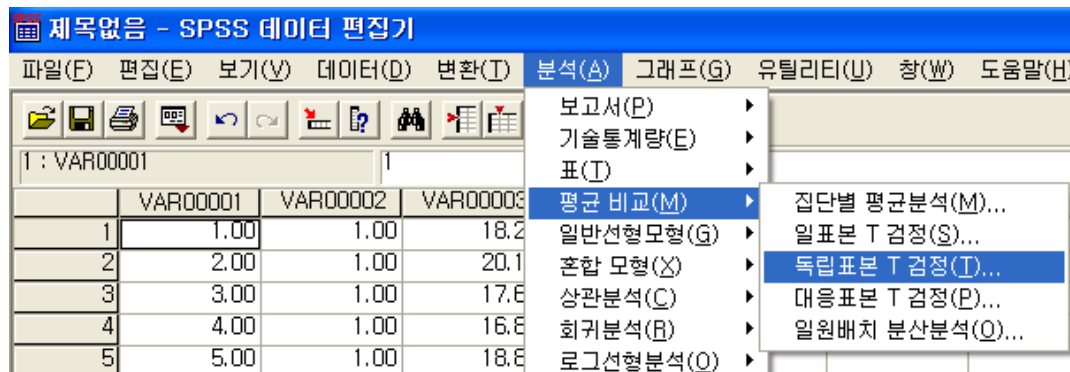
a. Lilliefors 유의확률 수정

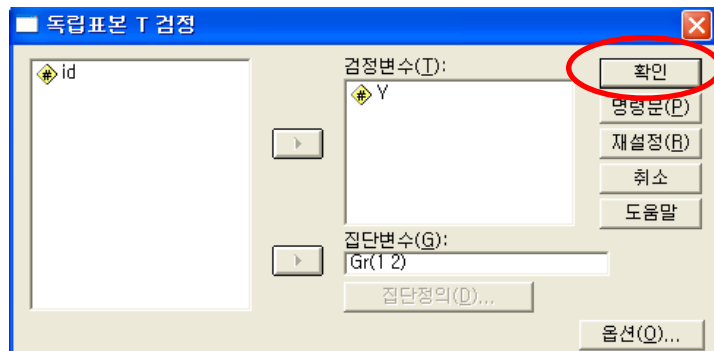
## - 정규성 검정의 가설

- 1) 귀무가설 : (각 집단별) 측정 자료는 정규분포를 따른다
- 2) 유의확률이 0.05보다 크면 유의수준 5%수준에서 정규분포를 따른다고 해석

## 2) 등분산 검정

- 두 집단의 평균 차이를 비교하기 위해 두 집단의 분산이 동일한지를 검정
- 귀무가설 : 두 집단의 모분산은 동일하다 (양측검정으로 수행)
- SPSS





독립표본 검정

		Levene의 등분산 검정		평균의 동일성에 대한 t-검정						
		F	유의확률	t	자유도	유의확률 (양쪽)	평균차	차이의 표 준오차	차이의 95% 신뢰구간	
Y	등분산이 가정됨	.001	.980	1,522	12	.154	.95714	.62879	하한	상한
	등분산이 가정되지 않음			1,522	11,996	.154	.95714	.62879	-.41287	2,32715
									-.41292	2,32721

## – 등분산 검정 방법

- Levene의 등분산 검정 결과를 해석
- 유의확률이 0.05보다 크면 유의수준 5%수준에서 귀무가설을 기각할 수 없으므로(채택하므로) 두 집단의 모분산은 동일하다는 귀무가설을 채택
- 따라서 두 집단의 모분산은 동일하다고 결론

### 3) T-검정

- 정규성 및 등분산 가정 만족 여부를 고려하여 t-검정 수행
- 귀무가설 : 두 집단의 평균은 동일하다/두 집단의 평균의 차이는 없다
- SPSS
  - 등분산 검정과 동일하게 수행

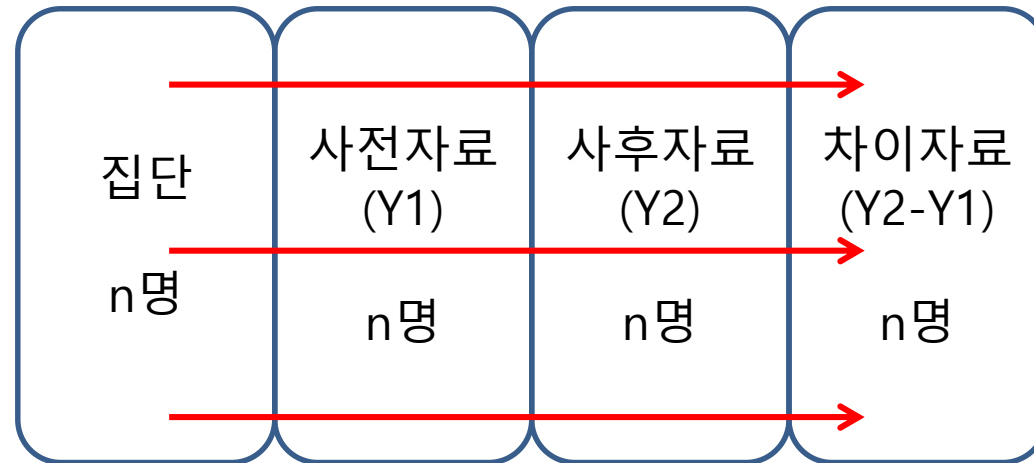
독립표본 검정

	Levene의 등분산 검정	평균의 동일성에 대한 t-검정							
		F	유의확률	t	자유도	유의확률 (양쪽)	평균차	차이의 표준오차	차이의 95% 신뢰구간 하한 상한
Y	등분산이 가정됨	.001	.980	1.522	12	.154	.95714	.62879	-.41287 2.32715
	등분산이 가정되지 않음			1.522	11.996	.154	.95714	.62879	-.41292 2.32721

- 등분산 가정을 만족하면, 등분산 가정된 칸을 이용하여 분석
- 등분산 가정을 만족하지 않으면, 등분산 가정되지 않음 칸을 해석
- 단측검정이면 유의확률/2로 판단
- 분석/해석 방법
  - 양측 : 집단1과 집단2의 평균 차이는 0.96이며, 두 집단의 평균 차이가 유의한지를 독립 t-검정으로 분석해 보면, 유의확률(p-값)이 0.154로 나타나 유의수준 5%에서 두 집단의 평균 차이가 없다는 귀무가설을 기각할 수 없으므로 두 집단의 평균 차이가 유의하지 않은 것으로 나타났다(유의하다면 집단1과 집단2의 평균 차이는 유의하다고 해석, 어느 집단이 높다는 결론은 내리지 않는 것으로 설명).
  - 단측 : 집단1과 집단2의 평균 차이는 0.96이며, 두 집단의 평균 차이가 유의한지를 독립 t-검정으로 분석해 보면, 유의확률(p-값)이 0.0777로 나타나 유의수준 5%에서 두 집단의 평균 차이가 없다는 귀무가설을 기각할 수 없으므로 두 집단의 평균 차이가 유의하지 않은 것으로 나타났다(만약 유의하다면 집단 1이 집단2보다 평균이 더 높은 것으로 나타났다고 설명 가능).

# 대응자료 t-검정

- 의미 : 동일 집단 대상으로 두 관심 변수의 평균 차이를 검정하는 방법
- 연구설계 : 자료수집



## – 가정

- 1) 동일 집단 대상으로 두 관심 변수의 자료를 측정
- 2) 연구 관심변수(Y)는 수치자료로 측정
- 3) 측정 자료는 정규분포로 가정 : 실제 연구에서는 분석을 통해 확인해야 함

- 정규성 검정
  - 검정 기준 : 택일 가능, 대체로 2)번으로 주로 확인
    - 1) 개별 자료( $Y_1, Y_2$ )에 대해 확인 가능
    - 2) 차이 자료  $d=Y_2-Y_1$ 에 대해 확인 가능
  - ✓ 정규성 검정 과정 : 독립집단 t-검정과 동일
- 등분산 가정
  - 두 자료에 대한 차이 자료를 이용하면 두 자료가 한 개의 관심변수 자료( $d$ )로 축소되므로 등분산 검정은 필요하지 않음



## • 결과 해석

대응표본 통계량

		평균	N	표준편차	평균의 표준오차
대응 1	V1	222,1250	8	25,91986	9,16406
	V2	224,5000	8	27,90801	9,86697

대응표본 상관계수

	N	상관계수	유의확률
대응 1 V1 & V2	8	.987	.000

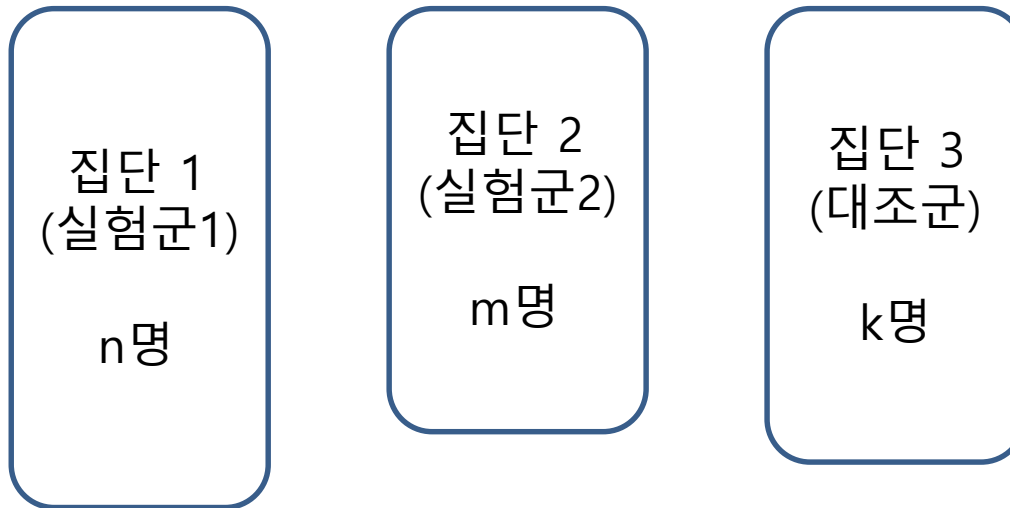
대응표본 검정

	대응차					t	자유도	유의확률 (양측)
	평균	표준편차	평균의 표 준오차	차이의 95% 신뢰구간				
				하한	상한			
대응 1 Y1 - Y2	-2,37500	4,83846	1,71065	-6,42006	1,67006	-1,388	7	,208

- 유의확률(p값)은 양측 검정이므로 단측검정이면 유의확률/2로 판단
- 양측 : 사전과 사후의 차이는 -2.375로 사후가 높지만 사전과 사후 차이가 유의하다는 연구가설(대립가설 관점)에 대한 대응자료 t-검정 결과는 p값이 0.208로 유의수준 5%보다 크므로 사전과 사후의 평균은 동일하다는 귀무가설을 기각할 수 없으므로 사전과 사후의 차이는 유의하지 않다고 해석
- 단측 : 사전과 사후의 차이는 -2.375로 사후가 높지만 사후가 사전보다 더 높다는 연구가설(대립가설 관점)에 대한 대응자료 t-검정 결과는 p값이 0.104로 유의수준 5%보다 크므로 사전과 사후의 평균은 동일하다는 귀무가설을 기각할 수 없으므로 사전과 사후의 차이는 유의하지 않다고 해석

# 분산분석

- 의미 : 세 집단의 평균 차이를 검정하는 방법
- 연구설계 : 자료수집



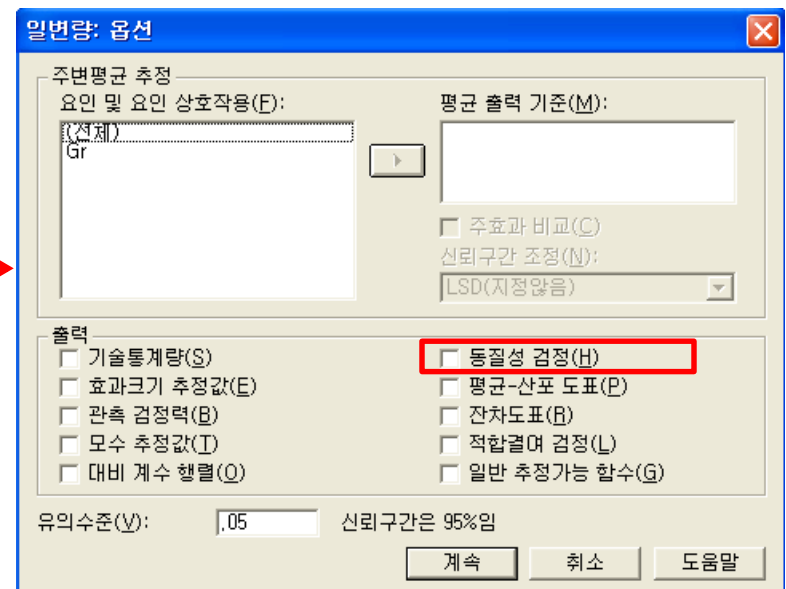
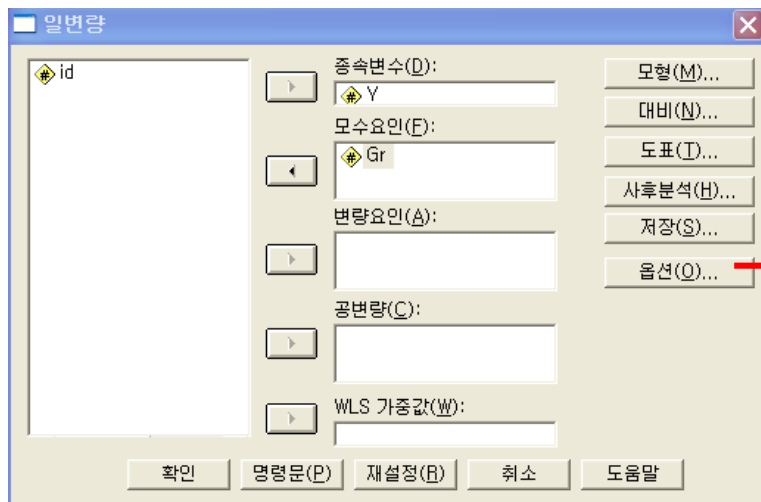
## – 가정

- 1) 세 집단은 서로 독립
- 2) 연구 관심변수(Y)는 수치자료로 측정
- 3) 측정 자료는 정규분포로 가정 : 실제 연구에서는 분석을 통해 확인해야 함

- 추가 가정의 검토
  - 정규성 검정
    - 각 집단별 측정 자료가 정규분포를 따르는지 확인
    - 독립진단 t-검정의 내용과 동일
    - 각 집단의 사례 수가 동일하다면 정규성 검정은 생략 가능
  - 등분산 가정
    - 세 집단의 모분산이 동일한지 확인/검정이 필요
    - SPSS에서 검정 가능 : “일반선형모형”을 이용하면 가능
  - 참고 : 집단내 대상간 독립성 가정의 검토도 가능

- [참고] 등분산 검정 : SPSS

- 귀무가설 : 각 집단의 모분산은 동일하다
- 분석 과정 : 분산분석 과정에서 검정 가능(예 ; 일반선형모형)



- 분석 결과

오차 분산의 동일성에 대한 Levene의 검정<sup>a</sup>

종속변수: Y

F	자유도1	자유도2	유의확률
.016	2	18	.984

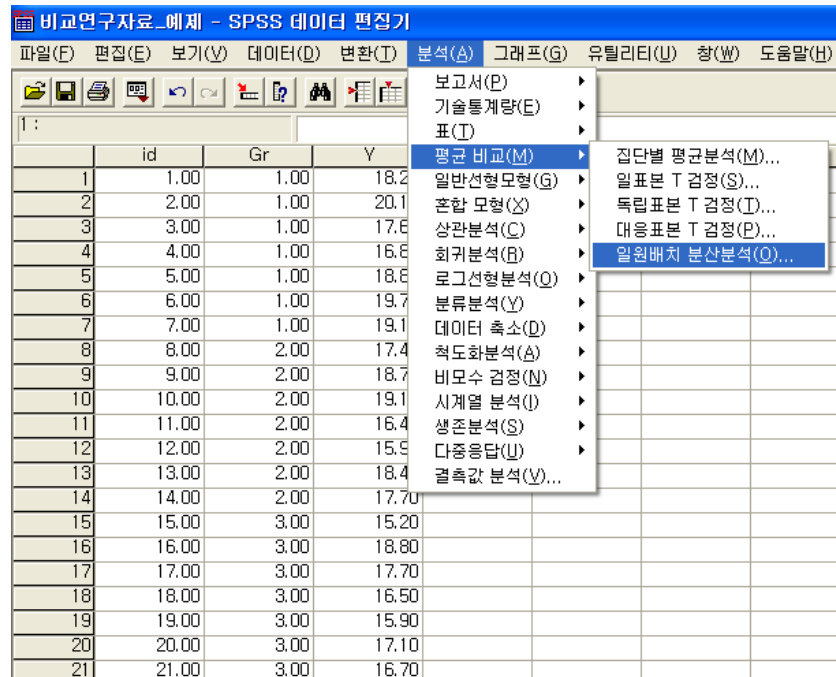
여러 집단에서 종속변수의 오차 분산이 동일한 영가설을 검정합니다.

a. 계획: Intercept+Gr

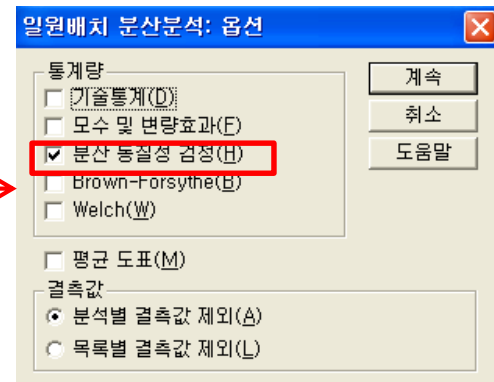
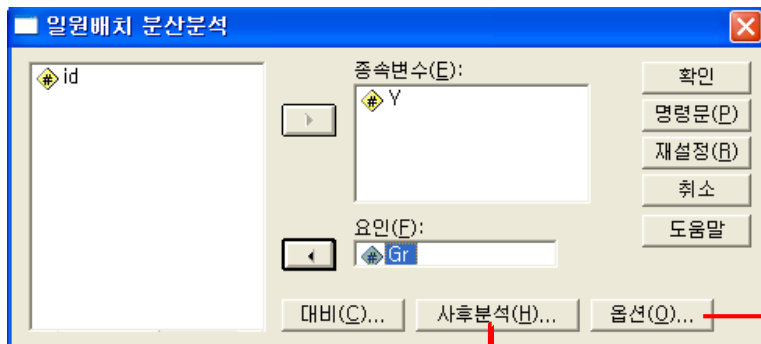
# SPSS : 일원배치 분산분석

## • 방법 1 : 일원배치 분산분석

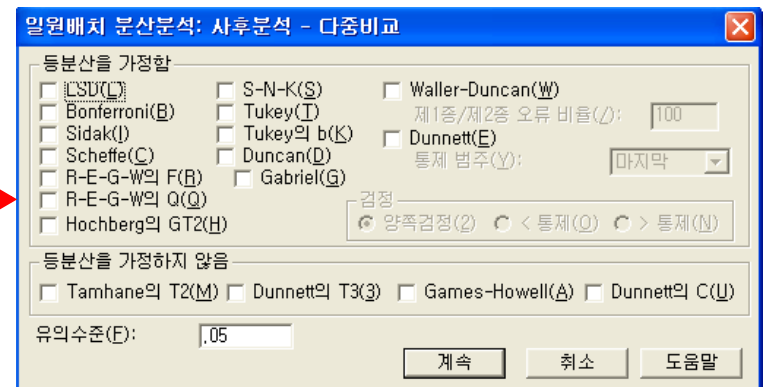
- 하나의 요인(하나의 집단/그룹 기준 변수)에 대한 분산분석만 가능
- 귀무가설 : 그룹별 평균은 동일하다/그룹별 효과의 차이는 없다
- 대립가설 : 최소한 하나의 집단은 평균이 다르다는 의미로 표현



## 분산동일성 검정



사후분석 : 분산분석에 대한 귀무가설이  
기각된 경우에만 추가 분석 수행



## • 일원배치 분산분석 결과

분산분석

Y	제곱합	자유도	평균제곱	F	유의확률
집단-간	11,007	2	5,503	3,968	.037
집단-내	24,963	18	1,387		
합계	35,970	20			

### – 사후분석 동일 집단군

Student-Newman-Keuls <sup>a</sup>			
Gr	N	유의수준 = .05에 대한 부집단	
		1	2
3,00	7	16,8429	
2,00	7	17,6571	17,6571
1,00	7		18,6143
유의확률		.212	.146

동일 집단군에 있는 집단에 대한 평균이 표시됩니다.

a. 조화평균 표본 크기 = 7,000을(를) 사용

귀무가설 기각인 경우만  
분석/설명

분석/설명 :

1) 세 집단의 평균이 동일하다는 귀무가설에 대한 분산분석 결과를 보면, p값이 0.037로 유의수준 5%보다 작으므로 귀무가설은 기각됨. 따라서 유의수준 5%에서 세 집단의 평균은 최소한 한 집단에서 차이를 보여주므로 유의수준 5%에서 집단간 평균의 차이가 유의하다고 설명.

2) 최소한 한 집단의 평균이 다르므로 구체적으로 어떤 집단의 평균이 차이를 살펴보기 위해 SNK 방법으로 사후분석(다중비교)을 실시함. 그 결과, 집단 3과 집단2는 평균이 같은 동일그룹으로 나타났으며, 집단2와 집단 1도 동일그룹으로 나타남(집단 2는 두 그룹과 동일하나 집단1과 3은 평균이 다른 집단으로 보임.).

# SPSS : 이원배치 분산분석

## • 방법2 : 일반선형모형

- 일반적인 상황(하나이상의 요인들로 이루어지는 실험)에 대한 분산분석 수행
- 둘 이상의 요인에 대한 실험에서 상호작용(교호작용)효과 분석이 가능

비교연구자료\_예제 - SPSS 데이터 편집기

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

1 :

	id	Gr1	GR2	Y	변수	변수	변수
1	1.00	1	1	18.20			
2	2.00	1	2	20.10			
3	3.00	1	3	17.60			
4	4.00	1	1	16.80			
5	5.00	1	2	18.80			
6	6.00	1	3	19.70			
7	7.00	1	1	19.10			
8	8.00	2	2	17.40			
9	9.00	2	3	18.70			
10	10.00	2	1	19.10			
11	11.00	2	2	16.40			
12	12.00	2	3	15.90			
13	13.00	2	1	18.40			
14	14.00	2	2	17.70			
15	15.00	3	3	15.20			
16	16.00	3	1	18.80			
17	17.00	3	2	17.70			
18	18.00	3	3	16.50			
19	19.00	3	1	15.90			
20	20.00	3	2	17.10			
21	21.00	3	3	16.70			

데이터(D) 변환(T) 분석(A) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

Gr1 GR2

Gr1	GR2	변수	변수
1			
1			
1			
1			
1			
1			
1			
2			
2			
2			
2			
2			
2			
2			

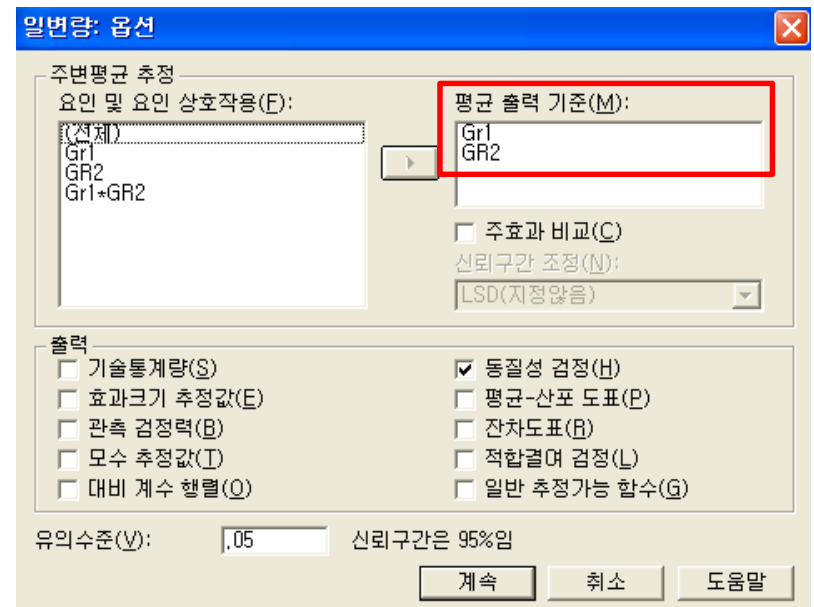
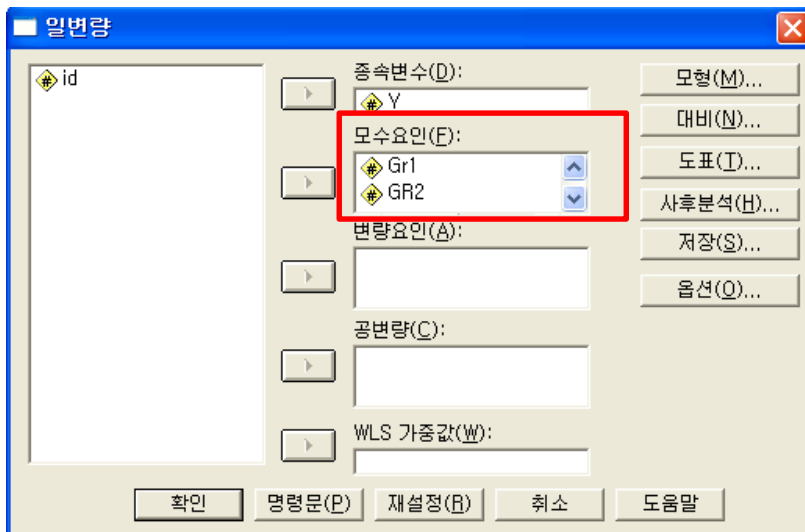
분석(A) 메뉴:

- 보고서(P)
- 기술통계량(E)
- 표(T)
- 평균 비교(M)
- 일반선형모형(G)**
  - 일반량(U)...**
  - 다변량(M)...
  - 반복측정(B)...
  - 분산성분 분석(V)...
- 혼합 모형(X)
- 상관분석(C)
- 회귀분석(R)
- 로그선형분석(Q)
- 분류분석(Y)
- 데이터 축소(D)
- 척도화분석(A)
- 비모수 검정(N)
- 시계열 분석(I)
- 생존분석(S)
- 다중응답(U)
- 결측값 분석(V)...



## • 가설

- 귀무가설 : 그룹간 평균은 동일하다/그룹간 평균의 차이는 없다
  - 그룹간 효과는 동일하다/효과의 차이는 없다
  - 상호작용효과의 검정 : 상호작용효과는 존재하지 않는다
- 대립가설 : 최소한 하나의 집단 평균은 다를 것이다
  - 상호작용효과의 검정 : 상호작용효과는 존재할 것이다



# • 분석 과정 : 이원배치 분산분석

## – 집단별 평균

### 1. Gr1

종속변수: Y

Gr1	평균	표준오차	95% 신뢰구간	
			하한값	상한값
1	18,711	,452	17,727	19,696
2	17,739	,452	16,754	18,723
3	16,961	,452	15,977	17,946

### 2. GR2

종속변수: Y

GR2	평균	표준오차	95% 신뢰구간	
			하한값	상한값
1	18,044	,452	17,060	19,029
2	18,006	,452	17,021	18,990
3	17,361	,452	16,377	18,346

- 설명 : 변수1( GR1) 기준의 세 집단 평균은 집단1이 가장 높고, 집단 3이 가장 작음
- 변수2(그룹 구분 변수/요인) 기준에서는 집단 1이 가장 높고, 집단 3이 가장 낮으나 그 차이가 크지 않음

## – 분산동일성 검정

오차 분산의 동일성에 대한 Levene의 검정<sup>a</sup>

종속변수: Y

F	자유도1	자유도2	유의확률
4,613	8	12	,009

여러 집단에서 종속변수의 오차 분산이 동일한 영가설을 검정합니다.

a. 계획: Intercept+Gr1+GR2+Gr1 \* GR2

## – 분산분석 결과

계제-간 효과 검정

종속변수: Y

소스	제 III 유형 제곱합	자유도	평균제곱	F	유의확률
수정 모형	19,430 <sup>a</sup>	8	2,429	1,762	,181
절편	6418,680	1	6418,680	4656,842	,000
Gr1	10,378	2	5,189	3,765	,054
GR2	1,988	2	,994	,721	,506
Gr1 * GR2	6,741	4	1,685	1,223	,352
오차	16,540	12	1,378		
합계	6618,600	21			
수정 합계	35,970	20			

a. R 제곱 = ,540 (수정된 R 제곱 = ,234)

## – 분산분석결과 해석 및 설명

- 변수1( GR1)과 변수2(GR2)를 고려한 관심변수 Y에 대한 실험자료를 2원배치 분산분석한 결과를 보면,
  - 변수1과 변수2의 상호작용에 대한 효과의 분석 결과를 보면, 효과가 존재하지 않는다는 귀무가설에 대한 검정 결과는 p-값이 0.352로 유의수준 5%에서 유의하지 않은 것으로 나타났다(귀무가설을 채택한 결과에 대한 설명).
  - 변수1에 대한 주효과에 대해서는 세 집단의 평균이 동일하다는 귀무가설에 대한 p-값이 0.054로 유의수준 5%에서 유의하지 않으며,
  - 변수 2에 대한 주효과에 대해서는 세 집단의 평균이 동일하다는 귀무가설에 대한 p-값이 0.506로 유의수준 5%에서 유의하지 않았다.

## – 사후분석

Gr1

### 다중 비교

종속변수: Y  
Dunnett T3

(I) Gr1	(J) Gr1	평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한값	상한값
1	2	,9571	,62879	,377	-,7704	2,6847
	3	1,7714*	,62689	,043	,0492	3,4937
2	1	-,9571	,62879	,377	-2,6847	,7704
	3	,8143	,63272	,510	-,9240	2,5525
3	1	-1,7714*	,62689	,043	-3,4937	-,0492
	2	-,8143	,63272	,510	-2,5525	,9240

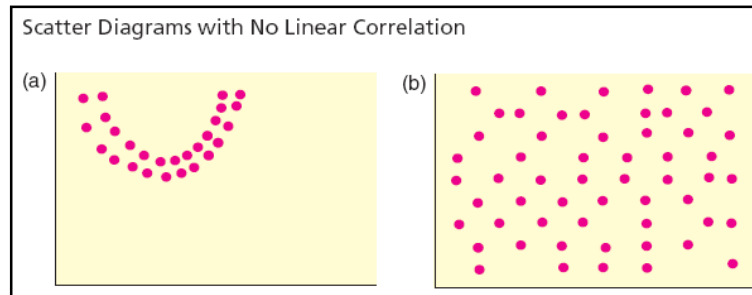
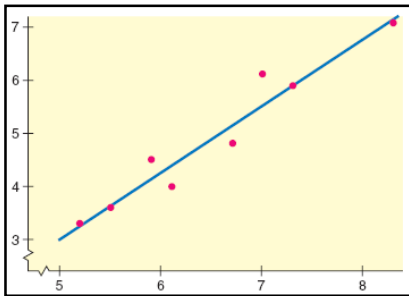
관측된 평균에 기초합니다.

\*, .05 수준에서 평균차는 유의합니다.

- 분산분석 검정에서 유의한 변수에 대해서는 사후분석을 실시하여 구체적으로 어느 집단의 평균이 다른지 확인하는 추가 검정을 실시하게 됨.
- 예 : 집단변수 1(Gr1)이 유의하다고 가정할 때,
  - 모분산이 동일하지 않으므로 동일하지 않을 때 사후분석하는 방법을 선택해 수행
  - Dunnett 방법으로 사후분석한 결과를 보면, 집단 1과 2는 동일한 평균을 가지며, 집단 3은 집단2와 동일한 평균을 가지는 것으로 나타나 집단 1과 2는 동일 집단, 집단2와 3이 동일집단으로 나타남.

# 상관분석

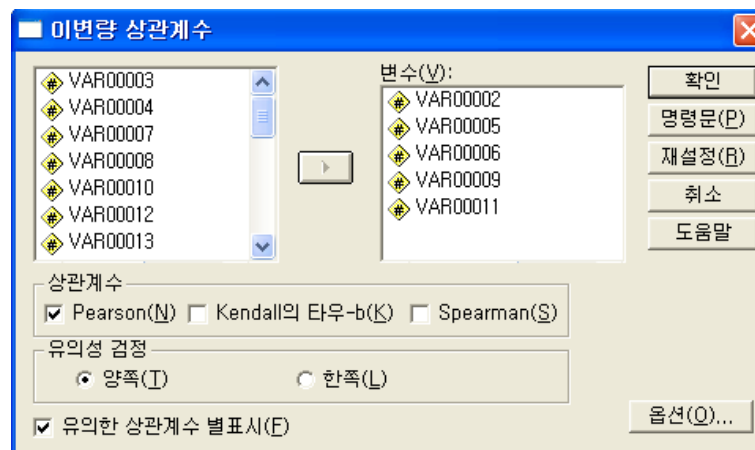
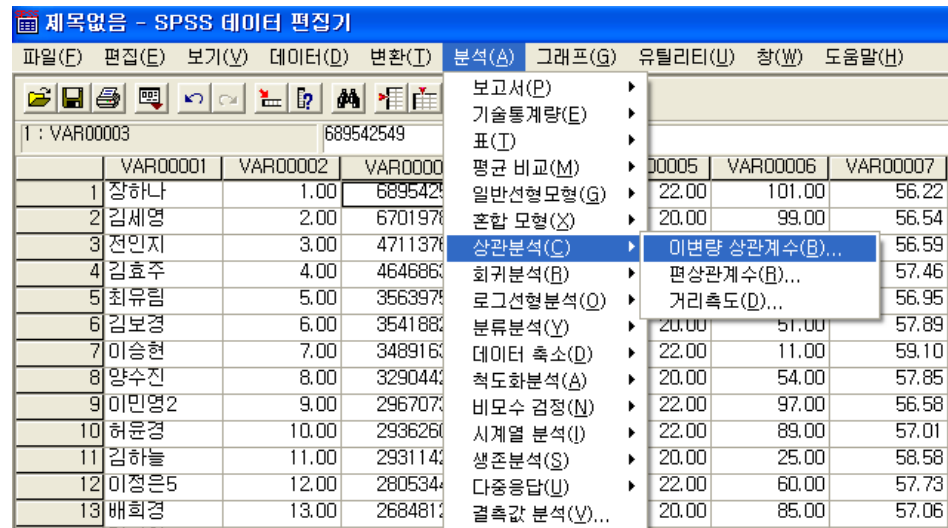
- 수치 자료의 (선형적) 관계를 분석하는 방법
  - 선형 관계를 제외한 관계분석은 불가능



- 가설
  - 귀무가설 : (선형적) 관계는 존재하지 않는다
  - 대립가설 : 관계(양의 관계, 음의 관계)는 존재한다
- 기본 검정 방법 : t-검정방법 이용
- 결과 해석
  - p값이 유의수준보다 작으면 귀무가설을 기각한다고 판단
  - 귀무가설을 기각하면 선형관계가 존재한다고 설명하고, 구체적인 관계는 대립가설을 보고 관계(양의 관계, 음의 관계)가 존재한다고 설명

# SPSS : 상관관계분석

- 분석 메뉴



- 분석 결과

상관계수

		VAR00002	VAR00005	VAR00006	VAR00009	VAR00011
VAR00002	Pearson 상관계수	1	-.685**	-.383**	.873**	-.855**
	유의확률 (양쪽)		.000	.000	.000	.000
	N	90	90	90	90	90
VAR00005	Pearson 상관계수	-.685**	1	.093	-.466**	.505**
	유의확률 (양쪽)	.000		.382	.000	.000
	N	90	90	90	90	90
VAR00006	Pearson 상관계수	-.383**	.093	1	-.379**	.378**
	유의확률 (양쪽)	.000	.382		.000	.000
	N	90	90	90	90	90
VAR00009	Pearson 상관계수	.873**	-.466**	-.379**	1	-.925**
	유의확률 (양쪽)	.000	.000	.000		.000
	N	90	90	90	90	90
VAR00011	Pearson 상관계수	-.855**	.505**	.378**	-.925**	1
	유의확률 (양쪽)	.000	.000	.000	.000	
	N	90	90	90	90	90

\*\* . 상관계수는 0.01 수준(양쪽)에서 유의합니다.

- 해석/설명

- 상금순위(var00002)와 참가대회수(var00005)의 상관계수는 -0.685(단측검정 p-값 0.000)로 유의수준 5%에서 귀무가설이 기각되어 유의하므로 상금순위가 높아지면 참가대회수가 적어짐을 알 수 있음.
- 상금순위(var00002)와 60타수획득순위(var00009)의 상관계수는 0.873(단측검정 p-값 0.000)로 유의수준 5%에서 귀무가설이 기각되어 유의하므로 상금순위가 높아지면 60타수 획득순위가 높아짐을 알 수 있음.

# 회귀분석

- 종속변수에 선형적인 영향을 미치는 독립변수들을 이용하여 회귀모형으로 1) 영향력의 크기, 2) 상대 영향력의 비교, 3) 종속변수의 결과 예측 등을 분석하는 통계적 기법
- 변수의 자료 형태
  - 종속변수 : 수치 자료
  - 독립변수 : 기본적으로 수치자료, 단, 비수치자료는 더미변수로 처리하여 사용
    - 더미변수(dummy variable) : (범주의 수-1)개의 더미변수(가변수) 생성

성별	d1
남자	1
여자	0

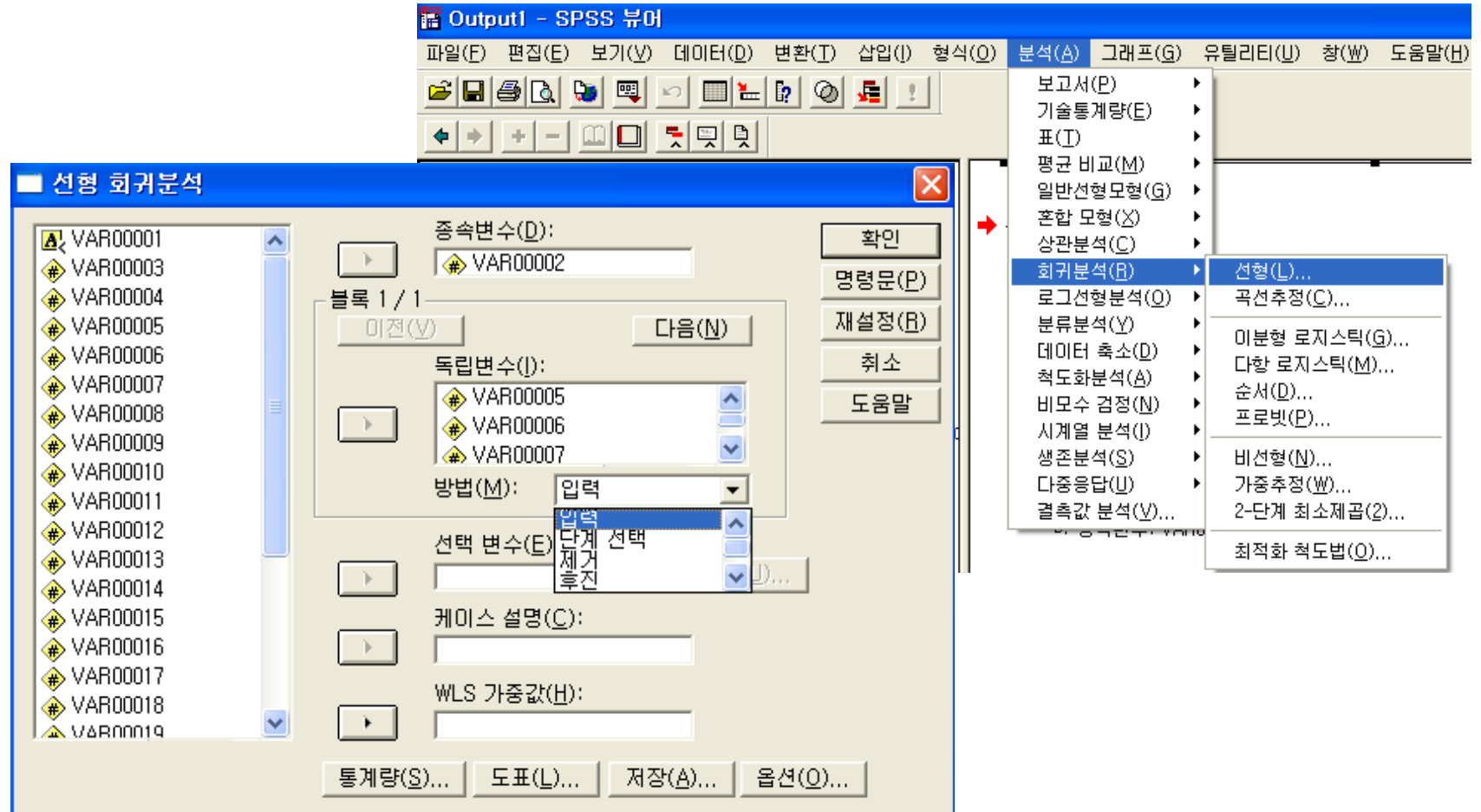
계절	d1	d2	d3
봄	1	0	0
여름	0	1	0
가을	0	0	1
겨울	0	0	0

- 모형 : 선형모형으로 가정



- 기본가정
  - 종속변수 혹은 오차항에 대한 가정이 요구
    - 정규성
    - 독립성
    - 등분산성
  - 선형성 : 독립변수와 종속변수의 관계에 대한 가정
  
- 회귀분석을 이용하는 기본 연구가설
  - 성별(독립변수)는 몸무게(종속변수)에 영향을 미칠 것이다
  - 계절은 운동량에 영향을 미칠 것이다
  - 퍼팅순위는 상금순위에 양의 영향을 미칠 것이다(순위가 일치한다는 의미)
  - 드라이빙 거리는 상금순위에 음의 영향을 미칠 것이다(거리가 멀면 순위가 감소함을 의미)

# SPSS : 회귀분석



## • 분석 과정

- 회귀모형의 유의성 검정 : 독립변수들이 종속변수를 회귀모형으로 설명하는지를 판단하는 검정(귀무가설 : 회귀모형은 유의하지 않다/독립변수들은 종속변수를 회귀모형으로 설명으로 설명할 수 없다)
  - 분산분석의 결과가 유의(귀무가설이 기각되는 상황)하게 나타나야 함

분산분석<sup>b</sup>

모형	제곱합	자유도	평균제곱	F	유의확률
1 선형회귀분석	67121,106	6	11186,851	208,406	,000 <sup>a</sup>
잔차	4455,294	83	53,678		
합계	71576,400	89			

a. 예측값: (상수), VAR00012, VAR00006, VAR00005, VAR00011, VAR00009, VAR00007

b. 종속변수: VAR00002

- 회귀모형의 설명력 설명
  - 결정계수(R제곱)를 이용하여 설명
  - 예 : 종속변수(var00002)에 대한 설명력을 보면, 독립변수들이 회귀모형을 통해 종속변수를 93.8% 설명하고 있음

모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	,968 <sup>a</sup>	,938	,933	7,32654

a. 예측값: (상수), VAR00012, VAR00006, VAR00005, VAR00011, VAR00009, VAR00007

- 회귀계수의 유의성 검정 : 개별적으로 독립변수가 종속변수에 미치는 영향력을 검정하는 과정
  - 귀무가설 : 독립변수는 종속변수에 영향을 주지 않는다
  - 대립가설 : 독립변수는 종속변수에 영향(양의 영향, 음의 영향)을 미친다

**계수<sup>a</sup>**

모형		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	-178.977	304.424		-.588	.558
	VAR00005	-2.345	.432	-.196	-5.431	.000
	VAR00006	.101	.145	.103	.698	.487
	VAR00007	3.950	5.132	.115	.770	.444
	VAR00009	.151	.091	.141	1.659	.101
	VAR00011	-.298	.472	-.048	-.632	.529
	VAR00012	.666	.071	.661	9.339	.000

a. 종속변수: VAR00002

- 유의확률(p값)을 이용하여 유의성을 검정(단측 검정은 p값을 1/2하여 검정)

## • 해석/설명

- 종속변수에 대한 회귀분석 결과를 보면, 유의수준 5%에서 var00005와 var00012만이 유의한 것으로 나타나 독립변수 중 var00005(참가대회수)와 var00012(타수순위)만이 종속변수에 미치는 영향이 유의한 것으로 나타남(종속변수에 유의하게 영향을 주는 것으로 나타남).
- var00005는 종속변수에 음의 영향을 미치므로 참가대회수가 적을수록 상금순위가 높아짐을 의미하고, var00012는 종속변수에 양의 영향을 미치므로 평균타수순위가 높을수록 상금순위도 높아짐을 의미함
- 영향을 미치는 두 변수의 상대적인 중요도를 비교하면, var00012의 표준화 회귀계수가 더 크므로 평균타수순위가 참가대회수보다 종속변수에 영향을 더 많이 미치고 있음

- 영향력의 크기 분석
  - 비표준화 회귀계수(B)를 이용하여 설명
- 상대적인 영향력의 크기 비교를 위한 방법
  - 표준화 회귀계수의 크기를 이용하여 비교
  - 부호는 방향성만을 의미하므로 절대값의 크기가 큰 변수가 영향력이 큼을 의미
- 예측을 위한 최종 모형의 선택
  - 유의한 독립변수만을 이용하여 최종 모형 선정
  - 변수 선택 방법으로 결정하는 방법
    - 전진선택법
    - 후진제거법
    - 단계별 선택법

- 단계별 회귀분석(stepwise regression)

- 전진선택 + 후진제거를 단계적으로 수행하여 유의한 독립변수들로 표현되는 최종 회귀모형의 결과를 제공하는 방법

The image shows the SPSS '선형 회귀분석' (Linear Regression) dialog box with the '단계 선택' (Stepwise Selection) method selected. The dependent variable is VAR00002, and the independent variables are VAR00005, VAR00006, and VAR00007. Below the dialog box is a table summarizing the models generated during the stepwise process.

모형 요약				
모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.954 <sup>a</sup>	.910	.909	8.57822
2	.965 <sup>b</sup>	.930	.929	7.56801
3	.968 <sup>c</sup>	.937	.935	7.24000

a. 예측값: (상수), VAR00012

b. 예측값: (상수), VAR00012, VAR00005

c. 예측값: (상수), VAR00012, VAR00005, VAR00009

## • 단계별 회귀분석 결과

분산분석<sup>d</sup>

모델	제곱합	자유도	평균제곱	F	유의확률
1 선형회귀분석	65100,850	1	65100,850	884,693	,000 <sup>a</sup>
잔차	6475,550	88	73,586		
합계	71576,400	89			
2 선형회귀분석	66593,499	2	33296,749	581,352	,000 <sup>b</sup>
잔차	4982,901	87	57,275		
합계	71576,400	89			
3 선형회귀분석	67068,481	3	22356,160	426,500	,000 <sup>c</sup>
잔차	4507,919	86	52,418		
합계	71576,400	89			

a. 예측값: (상수), VAR00012

b. 예측값: (상수), VAR00012, VAR00005

c. 예측값: (상수), VAR00012, VAR00005, VAR00009

d. 종속변수: VAR00002

계수<sup>a</sup>

모델		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	2,029	1,766		1,149	,254
	VAR00012	,960	,032	,954	29,744	,000
2	(상수)	47,944	9,128		5,252	,000
	VAR00012	,852	,036	,846	23,999	,000
	VAR00005	-2,154	,422	-,180	-5,105	,000
3	(상수)	51,329	8,805		5,830	,000
	VAR00012	,679	,067	,675	10,188	,000
	VAR00005	-2,372	,410	-,198	-5,784	,000
	VAR00009	,194	,064	,181	3,010	,003

a. 종속변수: VAR00002

- 상금순위(var00002)에 영향을 미치는 독립변수를 탐색하기 위해 단계별 회귀분석으로 분석한 결과, var00012(평균타수순위), var00005(참가대회수), var00009(60타획득순위)가 유의하게 영향을 미치는 것으로 나타남
- 유의하게 영향을 미치는 독립변수 중 평균타수순위가 상금순위에 영향을 가장 많이 미치고 있으며, 그리고 참가대회수, 60타 획득순위의 순으로 영향을 미침
- 최종 모형의 설명력은 결정계수가 93.8%로 설명력이 높게 나타남  
(설명력을 비교할 때는 수정결정계수를 사용하여 비교)